**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

**FACULTAD DE CONTADURÍA Y ADMINISTRACIÓN**

**PROGRAMACIÓN PARA LA EXTRACCIÓN DE DATOS(952)**

****

**PROYECTO FINAL**

**ARON ADRIAN MULATO TORRES**

**WEB SCRAPING**

**PROFESOR: JOSUE MIGUEL FLORES PARRA**

**TIJUANA, BAJA CALIFORNIA A 11 DE JUNIO DEL 2024.**

# ÍNDICE

[ÍNDICE 1](#_jsv64hje6ckf)

[Descripción del Problema 2](#_yynbkl1j8p4)

[Objetivos 2](#_ophl50ho11s4)

[RECOLECCIÓN DE DATOS 2](#_e1kyno87ohi)

[TRANSFORMACIÓN 5](#_u8fc0y5qzx5p)

[VISUALIZACIÓN 6](#_3brdh0rxr5w2)

[Tabla de datos de ligas 6](#_ayggbq8lj82i)

[Puntos por equipo en ligas seleccionadas. 7](#_l1pv181x9adu)

[Evolución de Puntos por Temporada en Ligas Seleccionadas 8](#_z8gqh06m5jbz)

[Distribución de Goles a Favor (GF) por Equipo en Ligas Seleccionadas 8](#_qde2instwyb)

[Goles a Favor vs Goles en Contra por Equipo 9](#_ks4kemxxy46a)

[Distribución de Puntos por Liga. 10](#_yu4gc4dvsnqt)

[Variabilidad de Puntos por Equipo en cada Liga 10](#_exfko3gcg4hb)

[CONCLUSIÓN 12](#_bq5sfl4g8awq)

[REFERENCIAS 13](#_2cio3chtf4lp)

#### **Descripción del Problema**

El objetivo de este proyecto es analizar y comparar el rendimiento de los equipos en las ligas de fútbol más importantes de Europa: La Liga (España), la Premier League (Inglaterra) y la Serie A (Italia). Buscamos identificar patrones y tendencias en el rendimiento de los equipos, jugadores destacados y factores que influyen en los resultados de los partidos.

#### **Objetivos**

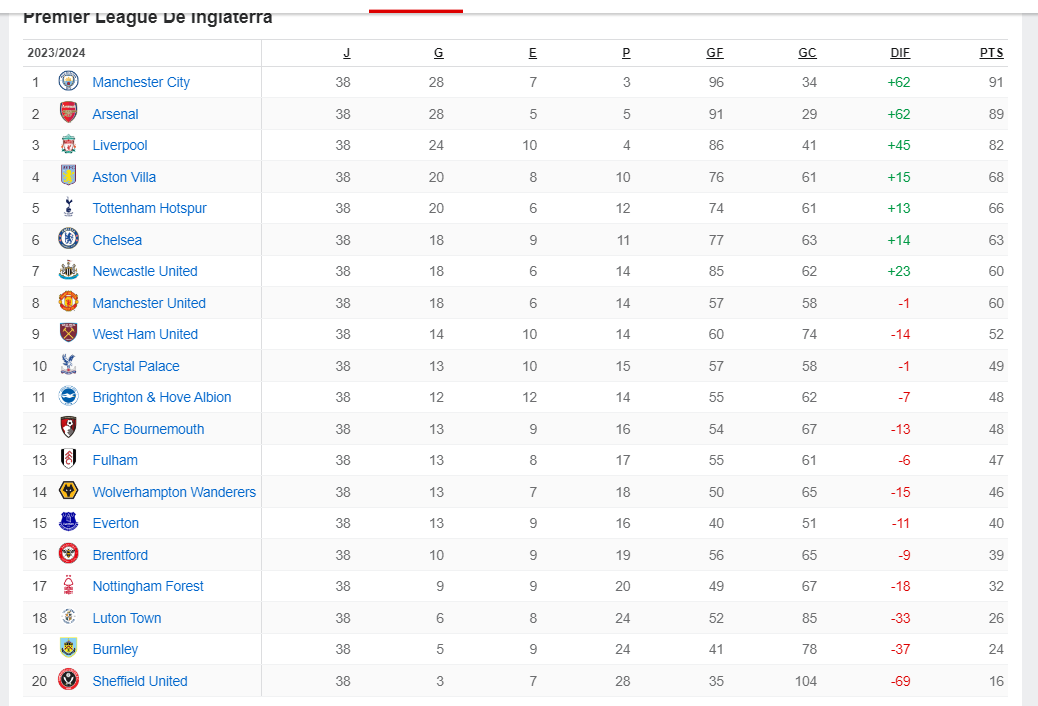
1. **Comparar el rendimiento de los equipos** en las tres ligas en términos de victorias, derrotas, empates, goles a favor y en contra.
2. **Analizar la evolución del rendimiento** de los equipos a lo largo de la temporada.
3. **Recolección de Datos** Extraer de manera efectiva los datos relevantes de la página de ESPN, centrándose en las puntuaciones de las tres ligas abarcando la información de la temporada 2003-2004 hasta la actualidad.
4. **Estructurar** los datos de manera clara y organizada, asegurándose de que la tabla incluya las 8 columnas esenciales y se muestran las estadísticas de estas ligas.
5. **Desarrollo de Dashboard** realizar un dashboard que presente la información recopilada de manera ordenada y fácilmente comprensible.

# RECOLECCIÓN DE DATOS

Recolectamos los datos de la página ESPN, los datos que tomamos en cuenta fueron las puntuaciones de la Serie A, LIga española y Premier league 2023-2024 en curso hasta la temporada 2003-2004

Hay 8 columnas que se conforman por Número de partidos jugados, partidos ganados, empates, derrotas, goles a favor, goles en contra, diferencia de puntos y puntos.

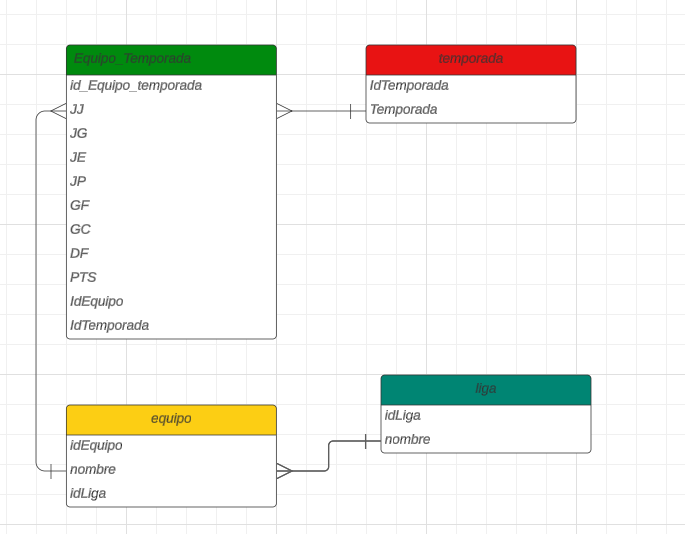






# TRANSFORMACIÓN

Modelo Relacional:



Diccionario de datos:

* Temporada: Año.
* Equipo: Nombre de equipo.
* JJ: Número de juegos jugados.
* JG: Número de juegos ganados.
* JE: Número de juegos empatados.
* JP: Número de juegos perdidos.
* GF: Número de goles a favor.
* GC: Número de goles en contra.
* DF: Diferencia de goles.
* PTS: Puntos obtenidos.
* idEquipo: Llave Primaria para tabla “Equipo”.
* idEquipo\_Temporada: Llave Primaria para tabla “Equipo\_Temporada”.
* idTemporada: Llave Primaria para tabla “Temporada”.
* idLiga: Llave Primaria para tabla “Liga”.

Nota: Nombre es para saber el nombre de la columna dependiendo la tabla.

# VISUALIZACIÓN



### **Tabla de datos de ligas**

Esta tabla presenta las estadísticas clave para los equipos en las ligas seleccionadas (La Liga, Serie A y Premier League) para varias temporadas. Los datos incluidos son:

* **Equipo**: Nombre del equipo.
* **Temporada**: Temporada específica en la que se jugaron los partidos.
* **JJ**: Partidos jugados.
* **JG**: Partidos ganados.
* **JE**: Partidos empatados.
* **JP**: Partidos perdidos.
* **GF**: Goles a favor.
* **GC**: Goles en contra.
* **DF**: Diferencia de goles (GF - GC).
* **PTS**: Puntos obtenidos.
* **Liga**: Nombre de la liga.

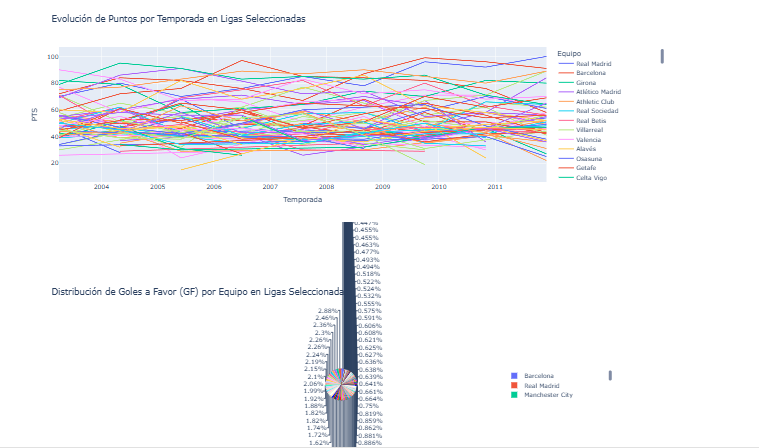
Por ejemplo, para el equipo Real Madrid en la temporada 2023-24, jugó 38 partidos, ganó 29, empató 8, perdió 1, marcó 87 goles a favor, recibió 26 goles en contra, tuvo una diferencia de goles de 61 y acumuló 95 puntos.

### **Puntos por equipo en ligas seleccionadas.**

Este gráfico de barras muestra la cantidad total de puntos acumulados por los equipos en las ligas seleccionadas (La Liga, Serie A y Premier League). Cada barra representa a un equipo y su altura corresponde a la cantidad de puntos obtenidos. Los colores indican a qué liga pertenece cada equipo:

* **Rojo**: Equipos de La Liga.
* **Azul**: Equipos de la Serie A.
* **Verde**: Equipos de la Premier League.

En este gráfico, se puede observar cómo equipos como el Real Madrid tienen una alta cantidad de puntos acumulados, comparado con otros equipos en las diferentes ligas.



### **Evolución de Puntos por Temporada en Ligas Seleccionadas**

Este gráfico de líneas muestra cómo han evolucionado los puntos de los equipos en las ligas seleccionadas a lo largo de las temporadas. Cada línea representa a un equipo, y la posición en el eje Y muestra los puntos acumulados en cada temporada.

* **Eje X (Temporada)**: Representa las temporadas desde 2004 hasta 2011.
* **Eje Y (PTS)**: Muestra los puntos acumulados por los equipos en cada temporada.
* **Líneas**: Cada línea representa a un equipo diferente. Por ejemplo, los equipos como Real Madrid, Barcelona y Atlético Madrid tienen sus propias líneas que indican su rendimiento a través de las temporadas.

Este dashboard permite observar la consistencia y las tendencias de rendimiento de los equipos a lo largo del tiempo. Por ejemplo, se puede ver cómo algunos equipos mantienen un rendimiento alto mientras que otros tienen más fluctuaciones en sus puntos.

### **Distribución de Goles a Favor (GF) por Equipo en Ligas Seleccionadas**

Este gráfico de pastel muestra la distribución de los goles a favor (GF) entre los equipos en las ligas seleccionadas. Cada segmento del pastel representa a un equipo y su tamaño indica el porcentaje de goles a favor que ha acumulado.

* **Segmentos del pastel**: Representan a los equipos y su contribución en términos de goles a favor. Por ejemplo, equipos como Barcelona, Real Madrid y Manchester City están resaltados con sus respectivos colores.
* **Colores**: Indican a qué liga pertenece cada equipo. En este caso:
  + **Azul**: Equipos de La Liga (por ejemplo, Barcelona y Real Madrid).
  + **Verde**: Equipos de la Premier League (por ejemplo, Manchester City).

Este dashboard proporciona una visión clara de cuáles equipos son los más efectivos ofensivamente en términos de goles a favor.

### 

### **Goles a Favor vs Goles en Contra por Equipo**

Este gráfico de dispersión muestra la relación entre los goles a favor (GF) y los goles en contra (GC) para los equipos en las ligas seleccionadas. Cada punto en el gráfico representa un equipo en una temporada específica.

* **Eje X (GF)**: Representa los goles a favor acumulados por los equipos.
* **Eje Y (GC)**: Representa los goles en contra recibidos por los equipos.
* **Puntos**: Cada punto representa a un equipo en una temporada específica. Los colores de los puntos indican a qué liga pertenece cada equipo:
  + **Azul**: La Liga.
  + **Rojo**: Serie A.
  + **Verde**: Premier League.

Este gráfico permite observar la relación entre la capacidad ofensiva y defensiva de los equipos. Por ejemplo, equipos con muchos goles a favor y pocos goles en contra estarán en la parte inferior derecha del gráfico, indicando un buen rendimiento tanto en ataque como en defensa.

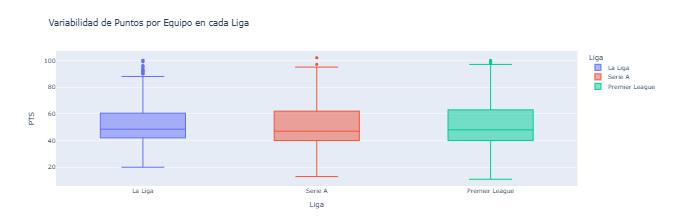
### 

### **Distribución de Puntos por Liga.**

Este histograma muestra la distribución de puntos obtenidos por los equipos en las ligas seleccionadas. Cada barra representa la cantidad de equipos que obtuvieron un rango específico de puntos en una temporada.

* **Eje X (PTS)**: Representa los puntos acumulados por los equipos.
* **Eje Y (Count)**: Representa la cantidad de equipos que obtuvieron esos puntos.
* **Barras**: Las barras están coloreadas según la liga a la que pertenecen los equipos:
  + **Azul**: La Liga.
  + **Rojo**: Serie A.
  + **Verde**: Premier League.

Este gráfico permite observar cómo se distribuyen los puntos en cada liga, identificando patrones como la cantidad de equipos que logran un rendimiento alto, medio o bajo en términos de puntos.



### **Variabilidad de Puntos por Equipo en cada Liga**

Este gráfico de caja y bigotes muestra la variabilidad de los puntos obtenidos por los equipos en cada una de las ligas seleccionadas (La Liga, Serie A y Premier League).

* **Eje X (Liga)**: Representa las ligas seleccionadas (La Liga, Serie A, Premier League).
* **Eje Y (PTS)**: Representa los puntos obtenidos por los equipos.
* **Cajas**: Las cajas representan el rango intercuartil (IQR), es decir, el rango entre el primer cuartil (Q1) y el tercer cuartil (Q3).
  + La línea dentro de la caja representa la mediana (Q2).
  + Los extremos de las cajas representan Q1 y Q3.
* **Bigotes**: Los bigotes se extienden hasta los puntos que no son considerados outliers. Los outliers son puntos que están a más de 1.5 veces el rango intercuartil desde Q1 o Q3.
* **Puntos individuales**: Los puntos que se encuentran fuera de los bigotes son considerados outliers y representan temporadas atípicas en términos de puntos obtenidos.

Cada liga tiene una caja de diferente color:

* **Azul**: La Liga.
* **Rojo**: Serie A.
* **Verde**: Premier League.

Este gráfico permite comparar la distribución y la variabilidad de los puntos obtenidos por los equipos en las diferentes ligas, destacando las diferencias en la competitividad y el rendimiento de los equipos.

# CONCLUSIÓN

El análisis de las principales ligas de fútbol europeas - La Liga, Premier League y Serie A - ha revelado información valiosa sobre el rendimiento de los equipos. Equipos como el Real Madrid han demostrado una consistencia notable en la obtención de puntos temporada tras temporada. A través del gráfico de líneas, se observó que algunos equipos mantienen un rendimiento constante mientras que otros presentan fluctuaciones significativas. El gráfico de pastel destacó la efectividad ofensiva de equipos como Barcelona y Manchester City, que se destacan por su capacidad para marcar goles. El gráfico de dispersión permitió identificar equipos que equilibran bien su capacidad ofensiva y defensiva. La distribución de puntos mostró cómo se reparten los puntos entre los equipos, con la Premier League y la Serie A destacándose por su competitividad diversa. Finalmente, el gráfico de caja y bigotes resaltó las diferencias en el rendimiento de los equipos, con algunas temporadas atípicas que reflejan un rendimiento excepcionalmente alto o bajo. Este análisis proporciona una visión clara y comparativa del rendimiento de los equipos en estas ligas, útil para entrenadores, analistas y aficionados del fútbol.

# REFERENCIAS

ESPN. (2023). *La Liga - Posiciones 2023-24*. Recuperado el 10 de junio de 2024, de<https://www.espn.com.mx/futbol/posiciones/_/liga/esp.1/temporada/2023>

ESPN. (2023). *Serie A - Posiciones 2023-24*. Recuperado el 10 de junio de 2024, de<https://www.espn.com.mx/futbol/posiciones/_/liga/ita.1/temporada/2023>

ESPN. (2023). *Premier League - Posiciones 2023-24*. Recuperado el 10 de junio de 2024, de<https://www.espn.com.mx/futbol/posiciones/_/liga/ENG.1/temporada/2023>